Advances in Medicine, Psychology and Public Health (AMPPH)

Original Article in Public Health

Predicting quality of life based on mental health state: A machine learning approach using Urban-HEART 2

Hamid SADEGHI[1], Masoud GHARIB[2], Vahidreza BORHANINEJAD[3], Vahid RASHEDI[4*]

[1]*Department of Electrical and Computer Engineering, Wayne State University, Michigan, Detroit, USA. Email: h.sadeghi.1991@gmail.com*
[2]*Orthopedic Research Center, Department of Rehabilitation Sciences, School of Allied Medical Sciences, Mazandaran University of Medical Sciences, Sari, Iran. Email: gharib_masoud@yahoo.com. ORCID: 0000-0002-6368-9736*
[3]*Social Determinants of Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. Email: borhani777@yahoo.com. ORCID: 0000-0002-4689-6741*
[4]*Iranian Research Center on Aging, Department of Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. Email: vahidrashedi@yahoo.com. ORCID: 0000-0002-3972-3789*

** Correspondence*

Cite this paper as: Sadeghi H, Gharib M, Borhaninejad V, Rashedi V. Predicting quality of life based in mental health state: A machine learning approach using Urban-HEART 2. Adv Med Psychol Public Health. 2024;1(3):133-142. Doi:10.5281/zenodo.10900463

Received: 15 January 2024
Revised: 25 February 2024
Accepted: 28 March 2024

Abstract

Introduction: Quality of life (QoL) is a complex and multifaceted concept often used as an indicator in evaluating public policy. This study aimed to predict QoL based on mental health state using a machine learning approach. The analysis was conducted using data from the Urban Health Equity Assessment and Response Tool (Urban-HEART 2) survey conducted in Tehran, Iran.

Methods: This secondary analysis utilized data from the second round of the Urban-HEART 2 survey, which included 117,839 participants. Various machine learning (ML) algorithms were employed, including Random Forest, Decision Tree, Support Vector Machine (SVM), Naive Bayes, and Logistic Regression. Additionally, an unsupervised learning method, specifically k-means clustering, was used.

Results: Following data preparation, the k-means clustering algorithm identified five clusters based on mental health features. ML algorithms were then utilized to predict each participant's QoL label through distinct scores. The top-performing ML algorithms based on high scores were found to be Random Forest (0.994), Decision Tree (0.991), SVM (0.990), Naive Bayes (0.935), and Logistic Regression (0.934), respectively.

Conclusions: By implementing k-means clustering, we identified distinct clusters based on mental health features and assigned labels to each participant accordingly. Machine learning models accurately predicted the QoL label for each participant. All models achieved high scores (above 0.93), indicating that mental health features can reliably predict QoL labels with high accuracy.

Take-home message: Metacognitive learning strategies, especially regulation, significantly impact nursing students' academic success. Integrating these strategies into curricula can enhance learning outcomes and benefit educators and students alike in nursing education.

Keywords: academic performance; metacognitive learning strategies; nursing students; planning, control and regulation.

INTRODUCTION

Quality of life (QoL) is complex and subjective, often used to evaluate public policies and outcomes in health and social care. While the literature identifies key QoL domains applicable to adults of all ages, the importance attributed to these domains can vary among different age groups [1]. Since the 1980s, QoL has become increasingly important as a patient-reported outcome in mental health services [2]. QoL theory provides a framework for conceptualizing people's mental health needs, describing services, and reporting program evaluations. Several research groups have advanced the conceptualization and measurement of QoL in mental health [3,4].

Criticism has been directed towards QoL measures, noting that they are often developed based on the perspective of mental health professionals rather than considering the perspectives of individuals with mental health issues and their perceptions of what is important for their own QoL [5]. Mental disorders were the second leading cause of disease burden in terms of years lived with disability (YLDs) and the sixth leading cause of disability-adjusted life-years (DALYs) in the world in 2017, posing a severe challenge to health systems, particularly in low-income and middle-income countries [6]. Mental health is recognized as one of the priority areas in health policies worldwide and has also been included in the Sustainable Development Goals (SDG) [7]. In mental health, QoL is commonly understood as an individual's personal assessment of different aspects of their life. These aspects may include physical health, family relationships, financial situation, and overall well-being [8].

In recent decades, Iran has experienced rapid demographic, societal, and economic changes. These changes have been accompanied by a decrease in the population growth rate, resulting in a shift in the country's age structure. Notably, individuals aged 30-39 years have emerged as the largest age group compared to other 10-year categories [9]. A study conducted by Sharifi et al. in 2015 revealed that the most common category of psychiatric disorders in Iran was any anxiety disorder, affecting approximately 15.6% of the population. Among specific disorders, major depressive disorder was found to be the most prevalent, affecting 12.7% of individuals. Generalized anxiety disorder followed at 5.2%, and obsessive-compulsive disorder at 5.1% [10]. A policy for more people attending mental health services to recover and have a good QoL necessitates appropriate outcome measures [5]. The high prevalence of mental illness and the need for effective mental health care, combined with recent advances in AI, has led to an increase in explorations of how the field of machine learning (ML) can assist in the detection, diagnosis, and treatment of mental health problems [11].

In recent years, there has been a growing interest in applying Machine Learning to mental health research. Le Glaz et al. conducted a study employing Machine Learning and Natural Language Processing (NLP) techniques for mental health analysis. Additionally, their study highlighted the potential use of the developed models in the broader mental health field [12]. Tate et al. successfully implemented a machine learning algorithm to predict the likelihood of persistent mental health issues in adolescents. Furthermore, they discussed exploring machine learning techniques that could outperform traditional logistic regression methods [13]. Another study by Alharahsheh et al. focused on predicting the likelihood of depression using supervised machine learning methods. The data for this study was collected by the Busara Center in Kenya [14]. Srividya et al. employed k-means clustering to identify meaningful clusters, which were further used to label the data. Subsequently, various machine learning models were utilized to predict the test data labels [15].

The primary objective of this research was to employ a machine learning approach using data from an extensive population-based survey (Urban-HEART 2) in Tehran, Iran, to predict the QoL based on mental health features. Notably, there is a lack of prior studies exploring these methodologies within this population. This study addressed the following

research questions: 1) Can the QoL label be predicted based on mental health features? 2) What strategies should be employed to enhance this prediction?

METHODS

Design and participants

This study involved a secondary data analysis from the second round of the Urban Health Equity Assessment and Response Tool (Urban HEART-2) survey. The survey included a substantial population of 117,839 participants. A multistage cluster random sampling method was employed, covering 368 neighborhoods across 22 districts of Tehran, the capital of Iran, in 2011. Detailed information regarding the study design and sampling procedures can be found elsewhere [16].

Urban HEART serves as a decision-support tool to identify and mitigate health inequities within cities. Its primary functions include: 1) Enhancing the understanding of unequal health determinants, risks, and outcomes experienced by individuals from diverse socioeconomic backgrounds within a city or across multiple cities, 2) Utilizing evidence-based approaches for advocating and planning interventions that promote health equity, 3) Encouraging participation in collaborative efforts across sectors to address health inequities, and 4) Applying a health equity perspective when making policy and resource allocation decisions [17].

Instruments

GHQ-28

The General Health Questionnaire-28 (GHQ-28) is a self-report questionnaire used as a screening tool for psychological well-being, which Goldberg developed in 1978 [18]. The GHQ-28 requests participants to indicate how their health, in general, has been over the past few weeks, using behavioral items with a 4-point scale showing the following frequencies of experience: "not at all", "no more than usual", "rather more than usual" and "much more than usual". The minimum score for GHQ-28 is 28, and the maximum is 112. Higher scores indicate higher levels of distress [19]. Noorbala *et al.* pointed out that the GHQ-28 is a valid and reliable psychiatric screening tool for the Iranian population [20].

Quality of Life Questionnaire: The scale comprises twelve concise statements regarding QoL, presented in a five-level Likert format (always, very often, sometimes, rarely, never). To ensure its validity and reliability, the questionnaire underwent rigorous evaluation by experts, including a panel of national experts from diverse disciplines, who reviewed its face and content validity [21].

Models

The Support Vector Machine (SVM) is a versatile linear model employed for regression and classification tasks, applicable to linear and nonlinear problems. It was first proposed by Vapnik. SVM has garnered significant attention among researchers as a compelling supervised learning approach for regression and classification. The algorithm's primary goal is maximizing the geometric margin between classifiers while minimizing the classification error [22,23].

Logistic Regression (LR) is a widely utilized method for estimating the probability of an event or a specific class. It finds application in binary classification tasks, such as image classification, where the outcomes are limited to binary values (zero or one). The core equation defining the principles of LR is as follows:

$$y = 1/(1 + e^{-m*x-b})$$

Here, "m," "b," and "x" represent constants and independent variables, respectively. In the context of this study, "x" corresponds to the feature set concerning [24].

Decision Tree is named as such because it partitions or breaks down the dataset into a hierarchical, tree-like structure. Each level of the tree represents a decision or split based on the input features, and the final leaf nodes of the tree correspond to the predicted outcomes or class labels. The tree-like structure makes it easy to interpret and visualize the decision-making process, hence the name "Decision Tree."

Random Forests is a nonlinear machine learning model commonly employed for classification tasks. It constitutes an ensemble of Decision Trees, where each tree is trained on a random subset of the data. The name "Random Forests" originates from using random subsets during training, and the collective predictions of the individual trees contribute to the final classification output. Compared to standalone Decision Trees, this ensemble approach enhances accuracy, robustness, and generalization, rendering Random Forests a popular choice for various classification problems.

Naive Bayes is a linear supervised learning classifier rooted in the Bayes Theorem. This technique operates under the assumption that the features are independent. The formula of this algorithm [25] is as follows:

$$P(y|x_1, \dots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, \dots, x_n)}$$

In this formula, "y" represents a variable, and "$x_1,\dots,x_n$" refers to the features associated with the variable "y." Naive Bayes is widely used for classification tasks, leveraging the Bayes Theorem to estimate the probability of class membership based on the given features. The independence assumption simplifies the computation and makes Naive Bayes a computationally efficient and effective classifier.

Clustering is an unsupervised method used to group a dataset into different clusters based on the similarity of their traits. This study subjected the dataset to k-mean clustering, dividing the data into distinct clusters.

Statistical analysis

Python programming language (version 3.7.4) and the scikit-learn module (version 0.22.1) were employed to predict the QoL labels. Several machine learning algorithms were utilized to perform the QoL label prediction. Scikit-learn is a Python module encompassing a diverse range of machine-learning algorithms for supervised and unsupervised learning [26].

Ethical aspects

The study was approved by the Ethics Committee of the Iran University of Medical Sciences (IUMS) in November 2010 and conducted following the principles of the Declaration of Helsinki, following relevant guidelines and regulations. Informed consent was obtained from all individual participants included in the study.

RESULTS

A total of 117,839 subjects participated in the study. The mean participants' age was 35.35 ± 19.46 years. The socio-demographic characteristics of the sample are illustrated in Table 1.

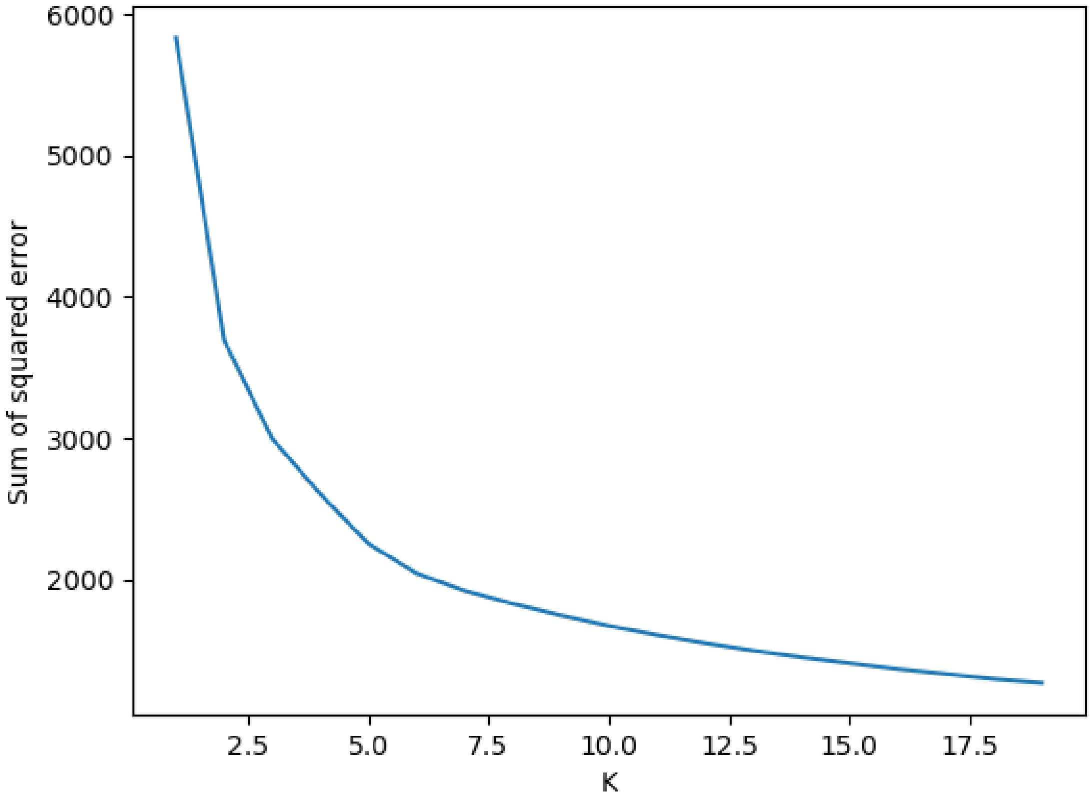
Table 1. Socio-demographic characteristics of the sample.

Variable		N = 117,839	
		N	%
Gender	Male	59,676	50.4
	Female	58,776	49.6
Marital Status	Married	68,294	57.6
	Widow	4,508	3.8
	Divorced	1,519	1.3
	Single	44,131	37.3
Schooling	No formal education	14,420	12.2
	1-5 years	13,689	11.6
	6-8 years	15,457	13.0
	9-12 years	45,630	38.5
	13-16 years	25,141	21.2

	≥ 17 years	4,415	3.5
Occupation	Employed	79,371	67.1
	Housekeeper	30,117	25.4
	Unemployed	8,964	7.5

First, k-mean clustering has been implemented to determine how many meaningful clusters exist (QoL labels). The elbow method was used to find the best cluster number. As shown in Figure 1, The best cluster number is five, in which the data labels are divided into five meaningful clusters (QoL labels). As shown in Figure 1, k=5, the number of errors has significantly reduced, which guarantees that the right number for k is five. These clusters have been used as QoL labels.

Figure 1. Elbow Method.



Five features (mental health features: QoL, somatization, anxiety, social dysfunction, and depression) and five labels (QoL labels) have been used for the dataset. The main aim is to predict labels (QoL labels) by having mental health features. The features used are as follows: Mental Health (QoL, Somatization, Anxiety, Social dysfunction, and depression).

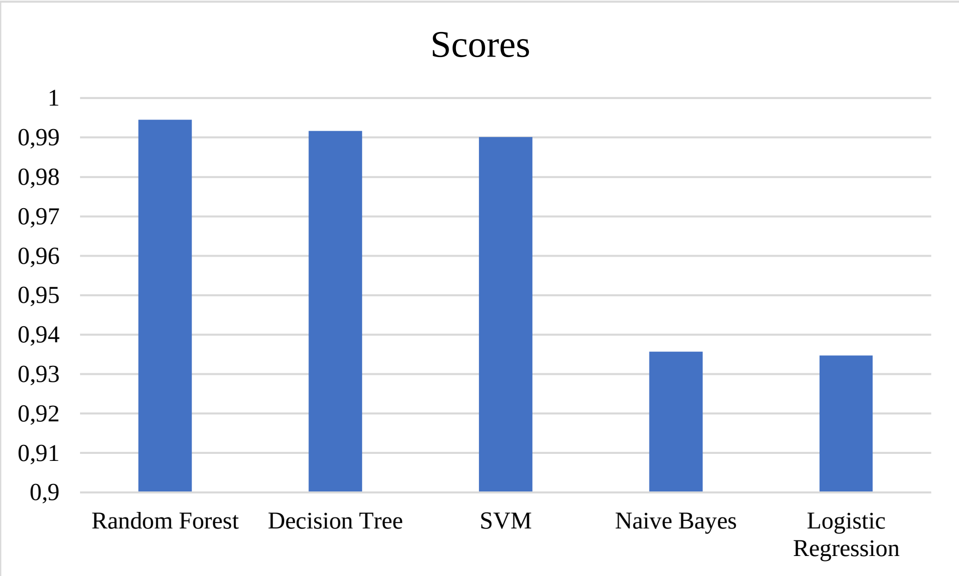
As shown in Table 2, the Random Forest (0.994) model has achieved the best machine learning score. In summary, the results demonstrated that most models used for the classification objective provided good scores. Interestingly, most models reached high scores (more than 0.93). Table 2 reveals that most classification models work great in prediction, whether linear classifiers such as SVM, Logistic Regression, and Naive Bayes or nonlinear classifiers such as Random Forest and Decision Tree. According to Table 2, there were no significant differences between linear and nonlinear models in terms of perdition, which means both linear and nonlinear models work well for classification purposes in this study. Model linearity does not affect prediction power in the current problem. QoL labels can be predicted based on mental health features. Seventy-five

percent of the dataset has been considered for training, and 25 percent for testing has been utilized. The scores were obtained after testing the data.

Table 2. Model scores.

Model Name	Score
Random Forest	0.9944
Decision Tree	0.9916
SVM	0.9900
Naive Bayes	0.9357
Logistic Regression	0.9347

Figure 2. Best Score.



As shown in Figure 2, the results demonstrate that the Random Forest gains the best score (0.994). Additionally, Decision Tree has achieved a high score (0.991) as a nonlinear machine learning classifier for classification problems. Similar to other machine learning models that have been used, SVM achieved a high score (0.990). As shown in Table 2, Naive Bayes has achieved almost similar scores in terms of prediction (0.935). Moreover, Logistic Regression has reached a good score (0.934).

DISCUSSION

This study used machine learning approaches to predict QoL labels based on mental health features. The analysis utilized data from a comprehensive population-based survey conducted in Tehran. This study implemented k-means clustering to identify appropriate clusters representing QoL labels.

The paper addresses diverse research investigating machine learning and artificial intelligence utilization in various fields. Chen Q, et al. established a system to recognize risk elements for heart disease [27]. Fraser KC, et al. employed mismatched machine-learning algorithms for diabetes detection [28]. Erickson et al. deployed machine learning on medical images to distinguish between benign and malignant tumors, effectively aiding radiologists in tumor identification [29]. The benefits of applying artificial intelligence in mental health research are evident. Tate et al. utilized machine learning techniques to

predict adolescent mental health [13], and Srividya et al. implemented clustering to predict outcomes based on mental health features [15]. Their findings demonstrate the potential of machine learning algorithms to predict QoL labels using mental health indicators, yielding high prediction scores (QoL labels based on mental health features).

In the domain of mental health research, NLP exhibits another compelling application. This is particularly evident in studies where NLP is employed to discern the potential risk of suicide among patients engaging in textual interactions with medical professionals. Within this context, NLP-based models emerge as a transformative tool, offering a cost-effective and streamlined avenue for identifying signs of suicide within text-based datasets. Through intricate linguistic analysis, these models bring forth an innovative means of pinpointing indicators of suicide risks, thereby aiding healthcare providers in their vigilant efforts to ensure the well-being of individuals. Underscored by its efficiency and affordability, this approach marks a significant stride toward enhancing mental health diagnostics and interventions [30].

Furthermore, contemporary research endeavors have proposed employing the Random Forest algorithm as a potent tool for uncovering response shift (RS) phenomena within QoL evaluations. This has garnered particular attention within the context of individuals grappling with the challenges of multiple sclerosis, where understanding the nuanced shifts in their perception of life quality stands paramount [31]. Through harnessing the capabilities of Random Forest, researchers aspire to delve deeper into the intricate interplay between disease impact and life satisfaction, shedding light on the multifaceted dynamics that underscore the QoL experience in these specific cases.

In the intricate landscape of patients grappling with neck and head cancer [32], decision tree algorithms emerge as indispensable tools that offer valuable insights into forecasting QoL labels. These algorithms leverage the intricate web of clinical variables, unraveling the complex interplay between disease progression, treatment approaches, and gender, ultimately offering a comprehensive understanding of the factors influencing QoL outcomes. Through the application of decision tree algorithms, researchers and healthcare practitioners can delve deeper into the nuanced relationships within this patient population, paving the way for more personalized interventions and improved QoL-enhancing strategies.

Exploring the realms of student well-being and academic excellence, researchers have delved into the application of SVM in the classification of students' QoL [33]. This insightful exploration highlights the significance of QoL and underscores its profound impact on academic performance. By harnessing SVM's robust classification capabilities, these studies have opened a window into a broader understanding of the intricate relationship between students' QoL and scholastic achievements. This research offers educators, institutions, and policymakers valuable insights into optimizing the academic journey by prioritizing and enhancing students' overall well-being.

In the realm of enhancing QoL, an additional layer of sophistication is introduced through the strategic utilization of Naive Bayes and Decision Tree methods. These cutting-edge techniques have been harnessed to predict and analyze air quality, a crucial factor that substantially influences QoL [34]. By intricately examining the interplay between air quality and well-being, these methodologies provide a comprehensive and multidimensional perspective on the factors contributing to an improved QoL. This research ventures beyond the surface, offering valuable insights into the potential mechanisms that could be manipulated to enhance air quality, thereby indirectly enhancing the overall QoL of individuals and communities. As such, these findings possess the potential to reshape urban planning, public health strategies, and environmental interventions to create a more conducive and nurturing environment for enhanced well-being.

Moreover, a meticulously crafted logistic regression model has emerged to enhance our understanding of post-transplant outcomes. This model, a result of meticulous research and analytical finesse, stands poised to forecast the trajectory of post-transplant health-related quality of life (HRQoL) in individuals [35]. By intricately examining many variables and their interplay, this predictive model sheds light on the complex dynamics that govern HRQoL post-transplantation. Its significance lies in its potential to identify those at risk of encountering a diminished QoL after undergoing transplantation. This innovative approach advances our predictive capabilities and promises to facilitate targeted interventions and

personalized care strategies, ultimately enriching the post-transplant journey for each individual. Through the lens of this logistic regression model, the landscape of post-transplant care transforms into a realm of possibilities where early interventions can pave the way for improved well-being and a higher QoL.

Demonstrating its efficacy in predictive tasks, logistic regression, a widely recognized supervised machine learning model, has exhibited its effectiveness in forecasting QoL labels. Through its systematic analysis of relevant features and patterns, logistic regression emerges as a robust tool for discerning and anticipating the intricate dynamics that shape an individual's QoL. Its capacity to discern patterns and relationships within data renders it a valuable asset in predictive modeling, offering insights that can contribute to informed decision-making and targeted interventions. As a testament to its versatility, logistic regression finds applicability in various domains, including healthcare, where it aids in predicting and understanding QoL outcomes. This model is a testament to the power of supervised machine learning techniques in unraveling the complexities of QoL prediction.

To delve even deeper into this field, there is a pressing need for further research that delves into additional prediction algorithms, like the K-Nearest Neighbors (KNN) algorithm or the intricate deep learning algorithms. Moreover, it has been proposed that enhancing the predictive capabilities could be achieved by integrating text data and harnessing the power of NLP algorithms. These potential avenues promise to enrich the precision and scope of QoL predictions, broaden the spectrum of variables considered, and tap into the nuances embedded in textual information. This calls for a comprehensive exploration of cutting-edge techniques, bridging the gap between traditional methods and the emerging potential of advanced algorithms to advance our comprehension of QoL and its intricate associations.

CONCLUSIONS

k-mean clustering was employed to identify the optimal number of clusters based on mental health features, and labels were assigned to each participant to indicate their QoL category. Machine learning models were then utilized to accurately predict the corresponding QoL label for each participant. Notably, all models achieved high scores exceeding 0.93, indicating that QoL labels could be accurately predicted using mental health features. This highlights the potential for mental health indicators to serve as reliable predictors of QoL with a high level of accuracy.

Author Contributions: Conceptualization: HS; methodology: HS, M.Gh; software: VB; validation: VR ; formal analysis, HS; investigation: HS, MGh, VB, VR; resources: VR; data curation: VB; writing—original draft preparation: HS; writing—review and editing: MGh, VB, VR; visualization: VR; supervision: VR; project administration: VR. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki.

Acknowledgments: None

Conflicts of Interest: The authors declare no conflict of interest.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

References

1. Brown J, Bowling A, Flynn T, editors. Models of quality of life: A taxonomy, overview and systematic review of the literature. London: European Forum on Population Ageing Research; 2004.
2. Buitenweg DC, Bongers IL, van de Mheen D, van Oers HA, van Nieuwenhuizen C. Subjectively different but objectively the same? Three profiles of QoL in people with severe mental health problems. Qual Life Res. 2018;27(11):2965-2974.
3. Short JL. Predicting Mental Health Quality of Life in Policing: Officers and Civilians. J Police Crim Psychol. 2021;36(2):276-287.

4. Nakagawa S, Yonekura S, Kanazawa H, Nishikawa S, Kuniyoshi Y, editors. Estimation of Mental Health Quality of Life using Visual Information during Interaction with a Communication Agent. 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE; 2020.
5. Connell J, O'Cathain A, Brazier J. Measuring quality of life in mental health: are we asking the right questions? Soc Sci Med. 2014;120:12-20.
6. Sagar R, Dandona R, Gururaj G, Dhaliwal R, Singh A, Ferrari A, et al. The burden of mental disorders across the states of India: the Global Burden of Disease Study 1990–2017. Lancet Psychiatry. 2020;7(2):148-161.
7. Saxena S, Funk M, Chisholm D. World health assembly adopts comprehensive mental health action plan 2013–2020. Lancet. 2013;381(9882):1970-1971.
8. van Nieuwenhuizen C, Nijman H. Qualify of life of forensic psychiatric inpatients. Int J Forensic Ment Health. 2009;8(1):9-15.
9. Statistical-Centre-of-ran. Population and Housing Censuses 2016. Accessed: 15 January 2024. Available from: https://www.amar.org.ir/english/Population-and-Housing-Censuses.
10. Sharifi V, Amin-Esmaeili M, Hajebi A, Motevalian A, Radgoodarzi R, Hefazi M, et al. Twelve-month prevalence and correlates of psychiatric disorders in Iran: the Iranian Mental Health Survey, 2011. Arch Iran Med. 2015;18(2):76-84.
11. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans Comput-Hum Interact. 2020;27(5):1-53.
12. Le Glaz A, Haralambous Y, Kim-Dufor D-H, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: Systematic review. J Med Internet Res. 2021;23(5):e15708.
13. Tate AE, McCabe RC, Larsson H, Lundström S, Lichtenstein P, Kuja-Halkola R. Predicting mental health problems in adolescence using machine learning techniques. PloS One. 2020;15(4):e0230389.
14. Alharahsheh YE, Abdullah MA, editors. Predicting Individuals Mental Health Status in Kenya using Machine Learning Methods. 12th International Conference on Information and Communication Systems (ICICS), IEEE; 2021.
15. Srividya M, Mohanavalli S, Bhalaji N. Behavioral modeling for mental health using machine learning algorithms. J Med Syst. 2018;42(5):1-12.
16. Rashedi V, Asadi-Lari M, Foroughan M, Delbari A, Fadayevatan R. Mental health and pain in older adults: findings from Urban HEART-2. Community Ment Health J. 2017;53(6):719-724.
17. World-Health-Organization. Urban HEART : urban health equity assessment and response tool: user manual Kobe, Japan 2010. Accessed: 15 January 2024. Available from: https://apps.who.int/iris/handle/10665/79061.
18. Goldberg D. Manual of the general health questionnaire. Windsor, England: NFER Nelson; 1978.
19. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med. 1979;9(1):139-145.
20. Nourbala AA, Bagheri YS, Mohammad K. The validation of general health questionnaire-28 as a psychiatric screening tool. Hakim Res J. 2009;11(4):47-53.
21. Asadi-Lari M, Vaez-Mahdavi MR, Faghihzadeh S, Montazeri A, Farshad AA, Kalantari N, et al. The application of urban health equity assessment and response tool (Urban HEART) in Tehran concepts and framework. Med J Islam Repub Iran. 2010;24(3):175-185.
22. Vapnik V. The nature of statistical learning theory: Springer Science & Business Media; 1999.
23. Durgesh KS, Lekha B. Data classification using support vector machine. J Theor Appl Inf Technol. 2010;12(1):1-7.
24. Peng C-YJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. J Educ Res. 2002;96(1):3-14.
25. Naive Bayes. Accessed: 10 January 2024. Available from: https://scikit-learn.org/stable/modules/naive_bayes.html.
26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
27. Chen Q, Li H, Tang B, Wang X, Liu X, Liu Z, et al. An automatic system to identify heart disease risk factors in clinical texts over time. J Biomed Inform. 2015;58:S158-S163.
28. Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex. 2014;55:43-60.

29. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics. 2017;37(2):505-515.
30. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med. 2019;49(9):1426-1448.
31. Boucekine M, Loundou A, Baumstarck K, Minaya-Flores P, Pelletier J, Ghattas B, et al. Using the random forest method to detect a response shift in the quality of life of multiple sclerosis patients: a cohort study. BMC Med Res Methodol. 2013;13(1):1-8.
32. de Melo NB, de Macedo Bernardino Í, de Melo DP, Gomes DQC, Bento PM. Head and neck cancer, quality of life, and determinant factors: a novel approach using decision tree analysis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2018;126(6):486-493.
33. Zainordin R, AM FN. Classification of students based on quality of life and academic performance by using support vector machine. J Acad. 2018;6(1):45-52.
34. Gore RW, Deshpande DS, editors. An approach for classification of health risks based on air quality levels. 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), IEE; 2017.
35. Khedmat H, Karami G-R, Pourfarziani V, Assari S, Rezailashkajani M, Naghizadeh M, editors. A logistic regression model for predicting health-related quality of life in kidney transplant recipients. Transplantation Proceedings; Elsevier; 2007.